GEOPHYSICS, VOL. 75, NO. 3 (MAY-JUNE 2010); P. R37–R46, 11 FIGS.
10.1190/1.3379323

Which data residual norm for robust elastic frequency-domain full waveform inversion?

Romain Brossier[1], Stéphane Operto[2], and Jean Virieux[3]

ABSTRACT

Elastic full-waveform inversion is an ill-posed data-fitting procedure that is sensitive to noise, inaccuracies of the starting model, definition of multiparameter classes, and inaccurate modeling of wavefield amplitudes. We have investigated the performance of different minimization functionals as the least-squares norm ℓ_2, the least-absolute-values norm ℓ_1, and combinations of both (the Huber and so-called hybrid criteria) with reference to two noisy offshore (Valhall model) and onshore (overthrust model) synthetic data sets. The four minimization functionals were implemented in 2D elastic frequency-domain full-waveform inversion (FWI), where efficient multiscale strategies were designed by successive inversions of a few increasing frequencies. For the offshore and onshore case studies, the ℓ_1-norm provided the most reliable models for P- and S-wave velocities (V_P and V_S), even when strongly decimated data sets that correspond to few frequencies were used in the inversion and when outliers polluted the data. The ℓ_2-norm can provide reliable results in the presence of uniform white noise for V_P and V_S if the data redundancy is increased by refining the frequency sampling interval in the inversion at the expense of computational efficiency. The ℓ_1-norm and the Huber and hybrid criteria, unlike the ℓ_2-norm, allow for successful imaging of the V_S model from noisy data in a soft-seabed environment, where the P-to-S-waves have a small footprint in the data. However, the Huber and hybrid criteria are sensitive to a threshold criterion that controls the transition between the criteria and that requires tedious trial-and-error investigations for reliable estimation. The ℓ_1-norm provides a robust alternative to the ℓ_2-norm for inverting decimated data sets in the framework of efficient frequency-domain FWI.

INTRODUCTION

Quantitative imaging of the earth's subsurface is essential for characterizing reservoirs, monitoring CO_2 sequestration, and evaluating soil properties in civil engineering applications. Full-waveform inversion (FWI) is a data-fitting procedure used to derive high-resolution quantitative models of the subsurface by exploiting the full information content of the data (Tarantola, 1984). When applied in the frequency domain, computationally efficient FWI algorithms (known as efficient frequency-domain FWI in the following) can be designed by exploiting the redundancy of the wavenumber coverage provided by wide-aperture acquisition surveys (Pratt and Worthington, 1990; Pratt, 1999). Decimation of the wavenumber redundancy can be implemented by limiting the inversion to a few judiciously chosen discrete frequencies at the expense of the signal-to-noise ratio (S/N) of the reconstructed models (Sirgue and Pratt, 2004; Brenders and Pratt, 2007a).

FWI potentially provides high-resolution models of the subsurface, but it suffers from two main difficulties. The first is related to the computational cost of the forward problem, the numerical resolution of the two-way wave equation in heterogeneous media for multiple sources. In the frequency domain, computationally efficient approaches based on direct solvers have been developed for 2D acoustic and elastic wave propagation (Jo et al., 1996; Stekl and Pratt, 1998; Hustedt et al., 2004; Brossier et al., 2009a). For 3D problems, the computational burden of the forward problem has motivated many efforts to develop efficient 3D modeling engines for frequency-domain FWI. These have been based on direct solvers, iterative solvers, hybrid direct/iterative solvers, and time-domain approaches (Nihei and Li, 2007; Operto et al., 2007; Plessix, 2007; Warner et al., 2007; Sirgue et al., 2008; Sourbier et al., 2008; Etienne et al., 2009).

The second difficulty is related to the ill-posedness and the nonlin-

Manuscript received by the Editor 24 July 2009; revised manuscript received 13 October 2009; published online 21 April 2010.
[1]Université Nice-Sophia Antipolis, Géoazur, Observatoire de la Côte d'Azur, Valbonne, France. E-mail: brossier@geoazur.unice.fr.
[2]Université Nice-Sophia Antipolis, Géoazur, Observatoire de la Côte d'Azur, Villefranche-sur-mer, France. E-mail: operto@geoazur.obs-vlfr.fr.
[3]Université Joseph Fourier, Laboratoire de Géophysique Interne et Tectonophysique, Grenoble, France. E-mail: jean.virieux@obs.ujf-grenoble.fr.

earity of the inverse problem, which is generally formulated as a least-squares local optimization so as to manage the computational cost of the forward problem (Tarantola and Valette, 1982). The ill-posedness of FWI mainly arises from the lack of low frequencies in the source bandwidth and the incomplete illumination of the subsurface provided by conventional seismic surveys. Consequently, the problem is highly nonlinear and the results strongly depend on the accuracy of the starting model in the framework of local optimization and on the presence of noise. Several hierarchical multiscale strategies that proceed from low frequencies to higher frequencies have been proposed to mitigate the nonlinearity of the inverse problem (Pratt and Worthington, 1990; Bunks et al., 1995; Sirgue and Pratt, 2004; Brossier et al., 2009a).

The noise footprint in seismic imaging conventionally is mitigated by stacking highly redundant multifold data. However, improving our understanding of inversion sensitivity to noise is a key issue, particularly when the data redundancy is decimated in the framework of efficient frequency-domain FWI. The least-squares objective function remains the most commonly used criterion in FWI, although it theoretically suffers from poor robustness in the presence of large isolated and non-Gaussian errors. Other norms can therefore be considered.

The least-absolute-values norm ℓ_1 is not based on Gaussian statistics in the data space; it was introduced into time-domain FWI by Tarantola (1987) and Crase et al. (1990) and is weakly sensitive to noise. Djikpéssé and Tarantola (1999) use the ℓ_1-norm successfully to invert field data from the Gulf of Mexico with time-domain FWI. Surprisingly, this norm has been used marginally during recent applications of FWI. Pyun et al. (2009) use an ℓ_1-like norm for frequency-domain FWI; the ℓ_1-norm is applied independently to the real and imaginary parts of the complex-valued wavefield. The resulting functional does not rigorously define a norm from a mathematical viewpoint because the functional does not satisfy the scalar multiplication property of norms ($\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$ for complex-valued scalar α and vector $\mathbf{x}$). The violation of the norm property makes the value of the functional vary with the phase of the residuals when the residual amplitude is kept constant. Despite this mathematical approximation, quite robust results have been obtained.

Alternative functionals, such as the Huber criterion (Huber, 1973; Guitton and Symes, 2003) and the hybrid ℓ_1/ℓ_2 criterion (Bube and Langan, 1997) can also be considered. Ha et al. (2009) apply the Huber criterion for frequency-domain FWI and illustrate its robust behavior compared to the ℓ_2-norm when considering a dense frequency sampling in inversion. All of these criteria behave as the ℓ_2-norm for small residuals and as the ℓ_1-norm for large residuals, thereby overcoming the nonderivability issue of the ℓ_1-norm for null residuals. A threshold, which needs to be defined, controls where the transition between these two different behaviors takes place with a more-or-less smooth shape, depending on the criteria. These hybrid criteria are efficient for dealing with outliers in data. However, they assume Gaussian statistics as soon as the ℓ_2-norm is used, leading to the difficult issue of estimating the threshold.

Our study presents applications of 2D elastic frequency-domain FWI for imaging realistic complex offshore and onshore structures in the presence of noisy synthetic data. Special emphasis is placed on the performance of different minimization criteria in the framework of efficient, elastic frequency-domain FWI.

In the next section, we briefly review the theoretical aspects of different possible norms and criteria that can be applied to frequency-domain FWI. Then we apply these objective functions to two synthetic data sets contaminated by ambient, random white noise with and without outliers. We assess the sensitivity of the inversion to noise in the case of decimated noisy data. The ℓ_2-norm is highly sensitive to non-Gaussian errors and requires consideration of denser frequency sampling to improve the S/N of the model. The ℓ_1-norm shows very robust behavior, even for highly decimated data, and therefore provides an interesting alternative to the ℓ_2-norm for efficient frequency-domain algorithms. Investigations of the Huber and hybrid criteria highlight the difficulties for finding the best threshold, which require tedious trial-and-error investigations.

THEORY AND ALGORITHM

Least-squares norm

The least-squares formalism provides the most common framework for frequency-domain FWI (Pratt, 1990; Pratt and Worthington, 1990). The ℓ_2 functional is usually written in the following form:

$$\mathcal{C}_{\ell_2}^{(k)} = \frac{1}{2}\Delta\mathbf{d}^{\dagger}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}, \tag{1}$$

where $\Delta\mathbf{d} = \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{calc}}^{(k)}$ is the data misfit vector for one source and one frequency, computed as the difference between the observed data $\mathbf{d}_{\text{obs}}$ and the modeled data $\mathbf{d}_{\text{calc}}^{(k)}$ in the model $\mathbf{m}^{(k)}$. The value k is the iteration number of the nonlinear iterative inversion. The dagger superscript indicates the adjoint operator, and $\mathbf{S}_d$ is a diagonal weighting matrix applied to the misfit vector to scale the relative contributions of each of its components.

Differentiating $\mathcal{C}_{\ell_2}^{(k)}$ with respect to the model parameters gives the following expression of the gradient:

$$\mathcal{G}_{\ell_2}^{(k)} = \mathcal{R}\{\mathbf{J}^t\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}^*\}, \tag{2}$$

where $\mathbf{J}$ is the Fréchet derivative matrix, t denotes the transpose operator, * denotes the conjugate operator, and $\mathcal{R}$ denotes the real part of a complex number. The gradient of the misfit function (equation 2) can be computed efficiently without explicitly forming $\mathbf{J}$ using the adjoint-state method (Plessix, 2006). This gives the expression

$$\mathcal{G}_{m_i\ell_2}^{(k)} = \mathcal{R}\left\{\mathbf{v}^t\frac{\partial\mathbf{A}^t}{\partial m_i}\mathbf{A}^{-1}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}^*\right\}, \tag{3}$$

where $\mathbf{A}$ is the forward-problem operator, which linearly relates the source $\mathbf{s}$ to the wavefield $\mathbf{v}$: $\mathbf{A}\mathbf{v} = \mathbf{s}$. The modeled seismic data used in FWI, $\mathbf{d}_{\text{calc}}$, are related to the seismic wavefield $\mathbf{v}$ by a projection operator that extracts the values of the seismic wavefield at the receiver positions. The sparse matrix $\partial\mathbf{A}/\partial m_i$ represents the radiation pattern of the diffraction by the model parameter m_i; hence, it gives some insight on the sensitivity of the data to a specific class of parameter as a function of the aperture angle. The gradient of the misfit function can be interpreted as a weighted zero-lag convolution between the incident wavefield $\mathbf{v}$ and the adjoint residual wavefield back propagated from the receiver positions $\mathbf{A}^{-1}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}^*$.

The misfit function and its gradient (equations 1 and 3) are given for one source and one frequency. For multiple sources and frequencies, the expressions are obtained by summing the contribution of each source and frequency.

Least-absolute-value norm

We can extend the ℓ_1-norm developed by Tarantola (1987) and Crase et al. (1990) for real arithmetic numbers to the complex arithmetic required by frequency-domain data through the misfit function:

$$\mathcal{C}^{(k)}_{\ell_1} = \sum_{i=1,N} |s_{d_i}\Delta d_i|, \tag{4}$$

where $|x| = (xx^*)^{1/2}$, N is the number of elements in the misfit vector for one source and one frequency, and s_{d_i} are the elements of the diagonal $\mathbf{S}_d$. The gradient of the misfit function is given by

$$\mathcal{G}^{(k)}_{\ell_1} = \mathcal{R}\{\mathbf{J}^t\mathbf{S}^t_d\mathbf{r}\} \text{ with } r_i = \frac{\Delta d_i^*}{|\Delta d_i|} \text{ for } 1 \le i \le N, \tag{5}$$

where we assume $|\Delta d_i| > 0$, considering the machine precision used. For all of the tests that we performed, we never met a case where $|\Delta d_i| = 0$. In the case of real arithmetic numbers, the term $\Delta d_i^*/|\Delta d_i|$ of expression 5 corresponds to the sign function (Tarantola, 1987; Crase et al., 1990).

Huber criterion

The Huber (1973) criterion can be defined for complex arithmetic numbers as

$$\mathcal{C}^{(k)}_{L_{\text{Huber}}} = \sum_{i=1,N} c_i \text{ with } \begin{cases} c_i = \dfrac{|s_{d_i}\Delta d_i|^2}{2\epsilon} & \text{for } |s_{d_i}\Delta d_i| \le \epsilon, \\ c_i = |s_{d_i}\Delta d_i| - \dfrac{\epsilon}{2} & \text{for } |s_{d_i}\Delta d_i| > \epsilon, \end{cases} \tag{6}$$

where ϵ is a threshold that controls the transition between the ℓ_1- and ℓ_2-norms. In equation 6, the Huber criterion is continuous for all Δd_i, particularly for the value that satisfies $|s_{d_i}\Delta d_i| = \epsilon$.

The gradient of the Huber functional is given by

$$\mathcal{G}^{(k)}_{L_{\text{Huber}}} = \begin{cases} \mathcal{R}\{\mathbf{J}^t\mathbf{S}^\dagger_d\mathbf{S}_d\mathbf{r}\} & \text{with } r_i = \dfrac{\Delta d_i^*}{\epsilon} \text{ for } |s_{d_i}\Delta d_i| \le \epsilon \text{ and } 1 \le i \le N, \\ \mathcal{R}\{\mathbf{J}^t\mathbf{S}^t_d\mathbf{r}\} & \text{with } r_i = \dfrac{\Delta d_i^*}{|\Delta d_i|} \text{ for } |s_{d_i}\Delta d_i| > \epsilon \text{ and } 1 \le i \le N. \end{cases} \tag{7}$$

Hybrid ℓ_1/ℓ_2 criterion

Bube and Langan (1997) introduce a hybrid ℓ_1/ℓ_2 criterion to overcome limitations of the Huber criterion that introduce artificial nonuniqueness in full-rank linear problems (Bube and Nemeth, 2007).

The hybrid ℓ_1/ℓ_2 functional can be written for complex arithmetic numbers as

$$\mathcal{C}^{(k)}_{L_{\text{hybrid}}} = \sum_{i=1,N} c_i \text{ with } c_i = \left(1 + \frac{|s_{d_i}\Delta d_i|^2}{\epsilon^2}\right)^{1/2} - 1, \tag{8}$$

where ϵ is the threshold between the ℓ_1 and ℓ_2 behaviors. The properties

$$c_i \approx \begin{cases} \dfrac{|s_{d_i}\Delta d_i|^2}{2\epsilon^2} & \text{for small } \Delta d_i \\ \dfrac{|s_{d_i}\Delta d_i|}{\epsilon} & \text{for large } \Delta d_i \end{cases} \tag{9}$$

show that the hybrid functional is asymptotically equivalent to the ℓ_2- and ℓ_1-norms for small- and large-amplitude residuals, respectively.

The gradient of the hybrid ℓ_1/ℓ_2 criterion is given by

$$\mathcal{G}^{(k)}_{L_{\text{hybrid}}} = \mathcal{R}\{\mathbf{J}^t\mathbf{S}^\dagger_d\mathbf{S}_d\mathbf{r}\} \text{ with } r_i = \frac{\Delta d_i^*}{\epsilon^2\left(1 + \dfrac{|s_{d_i}\Delta d_i|^2}{\epsilon^2}\right)^{1/2}} \text{ for } 1 \le i \le N. \tag{10}$$

Interpretation

Equations 2, 5, 7, and 10 clearly show that the gradients of the misfit functions have similar forms but different source terms for the back-propagated adjoint wavefield. This implies that the same FWI algorithm can be used to compute the gradients of the different misfit functions with the same computational cost, provided the source term of the adjoint back-propagated wavefield and the misfit function can be computed for each functional.

Figure 1 shows the misfit function and the source term of the back-propagated wavefield as functions of the real arithmetic unweighted misfit Δd for the four minimization criteria. The ℓ_2-norm naturally gives a high weight to large residual, which leads to a lack of robustness for this approach in the case of incoherent large errors in data. For the ℓ_1-norm, the data residuals are normalized according to their amplitudes, which gives clear insight into why this is expected to be less sensitive to large residuals. The Huber and hybrid criteria follow the ℓ_2 and ℓ_1 behaviors for small and large residuals, respectively, defined by the threshold ϵ with a different transition shape.

Algorithm

The 2D elastic frequency-domain FWI algorithm used in this study is described in Brossier et al. (2009a), where the reader is referred for a complete description of the algorithm. The algorithm embeds three main loops.

The outer loop is over the frequency groups, a set of frequencies inverted simultaneously. In the case of complex wave phenomena, such as P-to-S conversions, and multiples and surface waves, the simultaneous inversion of multiple frequencies better constrains the optimization for convergence toward the global minimum, taking into account more redundant information. The frequency interval within the groups is driven by the need to use as much redundant data

as possible during one FWI iteration yet keep the computations affordable. Moreover, the highest frequency of one frequency group is chosen by trial and error such that cycle-skipping artifacts are avoided.

The second loop is over time-damping factors that control the amount of information preserved over time in the seismograms during FWI. Time damping is applied in frequency-domain modeling by using complex-valued frequencies, given by $f_c = f + i\gamma$, where f is the real frequency and γ is the damping factor. This is equivalent to the damping of seismograms in time by $e^{-\gamma t}$ (Shin et al., 2002; Brenders and Pratt, 2007b). The damping can be applied from an arbitrary traveltime t_0, defined as the first-arrival traveltime. Moreover, this data preconditioning provides significant improvement over the S/N in the case of real data (Brenders et al., 2009). The choice of the damping factors is heuristic and is driven by the data complexity. We generally start inversion with a weak amount of data close to the first arrival using high γ values before progressively including later arrivals by decreasing the γ factor. Refer to Brossier et al. (2009a) for an illustration of the significant impact of the damping factors in elastic FWI of land data where body waves and surface waves are jointly inverted.

The third loop is over iterations of the nonlinear inversion of one frequency group and one damping factor. The two outer nested loops define two hierarchical multiscale levels in the inversion that mitigate the nonlinearity of the inversion, particularly in the case of elastic multiparameter inversion of complex data.

The forward problem is performed with a finite-element discontinuous Galerkin method for solving the elastodynamic equations in the frequency domain (Brossier et al., 2008). The linear system resulting from this discretization is solved in parallel using the MUMPS LU solver (Amestoy et al., 2006).

The optimization problem is solved with the quasi-Newton L-BFGS algorithm (Nocedal, 1980). In quasi-Newton algorithm, the updated model is given by

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \alpha^{(k)} H^{(k)} \mathcal{G}^{(k)}, \tag{11}$$

where $H^{(k)}$ is an approximation of the inverse of the Hessian, where the Hessian operator is formed by the second derivative of the misfit function with respect to the model parameters and $\alpha^{(k)}$ is the step length. The L-BFGS algorithm estimates the product $H^{(k)}\mathcal{G}^{(k)}$ at a negligible computational cost from few gradients and models of the previous iterations. Estimating the off-diagonal terms of the Hessian improves the imaging resolution by correctly scaling and deconvolving the gradient, compared to a more classic preconditioned conjugate gradient (Brossier et al., 2009a).

Newton or quasi-Newton algorithms usually are applied to quadratic or locally quadratic misfit functions. In contrast, the ℓ_1-norm and the Huber criterion are not twice continuously differentiable, a condition for the convergence of the L-BFGS algorithm. Nonetheless, the numerical examples performed in this study show that violating the convergence conditions did not significantly affect, in practice, the convergence of the L-BFGS algorithm. Similar conclusions are derived by Guitton and Symes (2003), who apply the L-BFGS algorithm with the Huber criterion for velocity analysis. The convergence of the L-BFGS algorithm for functionals that are not twice continuously differentiable can be interpreted by the fact that only a definite positive matrix approximation of the inverse of the Hessian — not the exact one — is computed in the L-BFGS algorithm from only the first derivative (i.e., gradient) of the misfit functions at previous iterations (Guitton and Symes, 2003). Moreover, the line search in the perturbation direction is performed by parabolic fitting to compute the optimal step length $\alpha^{(k)}$ and to guarantee the decrease of the misfit function, even for functionals that are not twice continuously differentiable.

For all applications presented in this study, the information of the five previous iterations are used for L-BFGS. The diagonal of the ℓ_2 pseudo-Hessian matrix (Shin et al., 2001) is used as an initial guess for the L-BFGS algorithm for each minimization criterion investigated. The aim of the diagonal pseudo-Hessian is to remove the geometric spreading signature of the incident and back-propagated wavefields from the gradient amplitudes.

The perturbation model estimated at each iteration is regularized with an adaptive Gaussian smoothing operator to filter out high-frequency artifacts that are not constrained by the current group of inverted frequencies. The local vertical correlation length of the Gaussian filter is defined according to the inverted frequency and the local wavespeeds, such that wavelengths significantly smaller than half the propagated wavelength (i.e., the maximum vertical resolution of FWI) are filtered out in the gradient (Sirgue and Pratt, 2004). The horizontal correlation length of the Gaussian smoother is estimated by trial and error and should be driven by the lateral heterogeneity of the structure as well as the dip illumination provided by the available source-receiver offsets.

The source is estimated in the FWI algorithm by a linear inversion (Pratt, 1999; his equation 17). For one frequency, this gives for the scalar source

$$s = \frac{\mathbf{g}^t \mathbf{d}^*_{\text{obs}}}{\mathbf{g}^t \mathbf{g}^*}, \tag{12}$$

where $\mathbf{g}$ denotes the incident Green's functions recorded at receiver positions, computed in the starting model of the current iteration. The source is estimated in the FWI algorithm with equation 12, once $\mathbf{g}$ are computed. The algorithm proceeds with computing the data residual $\Delta\mathbf{d}$, back propagating the residuals, computing the gradient and pseudo-Hessian, estimating the step length, and updating the model.

The alternative estimation of the source and the model is repeated at each iteration. Of note, the source is estimated with the least-squares norm even in the presence of outliers, whatever norm is

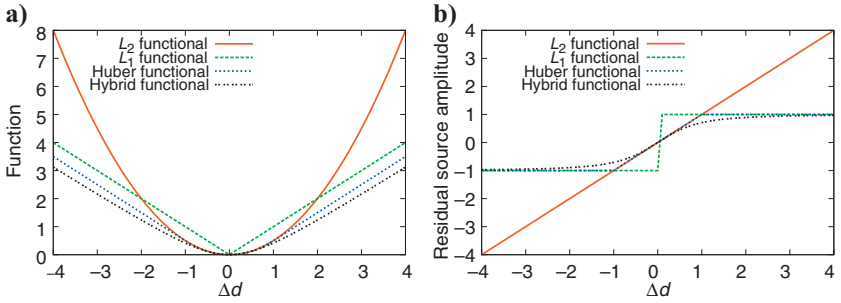

Figure 1. (a) The values of the criteria as functions of an unweighted real arithmetic misfit Δd. (b) The associated residual source amplitude in the gradient expression. ℓ_2, ℓ_1, and the Huber and hybrid functionals are shown in the red, green, blue, and black lines, respectively. The last two criteria are plotted for $\epsilon = 1$.

used for the model update. The source estimation was quite robust in the presence of noise for the examples shown hereafter: The error does not exceed 5% in amplitude and 1% in phase. The robustness of the source estimation can be attributed to the strong redundancy of the data for the monofrequency scalar source estimation.

For all numerical tests in this study, the threshold value ϵ for the Huber and hybrid criteria was fixed to $\epsilon = 0.2\ \text{mean}(|d_{\text{obs}_i}|)$. This value was shown practically to be less sensitive to outliers in the data than the one indicated by Guitton and Symes (2003), based on $\max(|d_{\text{obs}_i}|)$.

NUMERICAL TESTS: THE OFFSHORE VALHALL MODEL

Inversion setup

A first numerical example is based on the synthetic Valhall model (Figure 2), representative of oil and gas fields in shallow-water environments of the North Sea (Munns, 1985). The main targets are a gas cloud in the large sediment layer and, in a deeper part of the model, the trapped oil beneath the cap rock, which is formed of chalk. Gas clouds are easily identified by the low P-wave velocities, whereas their signature is much weaker in the V_S model. The selected acquisition mimics a four-component ocean-bottom cable (OBC) survey (Kommedal et al., 2004), with a line of 315 explosive sources positioned 5 m below the water surface and 315 three-component sensors on the seabed. This geologic setting leads to a particularly ill-posed problem for S-wave velocity reconstruction as a result of the relatively small shear-wave velocity contrast at the seabed, which prevents recording of significant P-to-S converted waves. A successful inversion requires a multistep hierarchical strategy in the manner of Sears et al. (2008) and as developed in Brossier et al. (2009b) for noise-free data. In this study, we assess the same approach for noisy data.

In the first step, the P-wave velocity is reconstructed from the hydrophone data. The forward problem is performed with the elastic discontinuous Galerkin method, but the V_S model is left unchanged during FWI. The aim of step 1 is to improve the V_P model so as to decrease the P-wave residuals significantly. During step 1, a coarse mesh adapted to the V_P wavelength is designed for computational efficiency. In this case, S-wave modeling is affected by numerical dispersion that does not significantly impact reconstruction of the V_P model. This first step is justified by the fact that the P-to-S converted waves have a minor footprint in the hydrophone component. This negligible sensitivity of the hydrophone data to the V_S structure allows for us the successful acoustic inversion of the elastic data computed in the Valhall model (Brossier et al., 2009b).

In step 2, the V_P and V_S models are reconstructed simultaneously from the horizontal and vertical components of the geophones. An amplification with a gain given by the power of two of the source/receiver offset is applied to the data through the matrix $\mathbf{S}_d$. This weighting increases the weight of the intermediate-to-long-offset data at which the converted P-to-S arrivals are recorded.

Five frequencies were inverted successively (2, 3, 4, 5, and 6 Hz). The starting frequency (2 Hz) is lower than the one available in the real OBC hydrophone data of Valhall (3.5 Hz) (Sirgue et al., 2009). However, a starting frequency as small as 2 Hz recently was used to perform acoustic FWI of ocean-bottom-seismometer data (Plessix, 2009). The use of such low starting frequency is required because V_S has a higher resolution power than V_P; the shorter propagated wavelength requires a lower starting frequency or a more accurate starting model. Our main concern in this study is to tune the elastic FWI with a reasonably realistic experimental setup such that differences in the behaviors of the different functionals are highlighted. In the following, we do not readdress the impact of starting frequency and starting model in FWI; rather, we focus on the comparative performances of different data residual functionals for a given starting-frequency/model pair.

During each frequency inversion, we used three time-damping factors ($\gamma = 2, 0.33, 0.1\ \text{s}^{-1}$) applied in cascade to the monochromatic data. For the smaller damping factor, the entire wavefield, including converted waves and free-surface multiples, was involved in the inversion. Starting models were built by smoothing the true models with a Gaussian filter, the vertical correlation length of which increased linearly from 25 to 1000 m with depth; the horizontal correlation length was fixed at 500 m (Figure 3). This smoothing reasonably mimicked the spatial resolution of a velocity model developed by refraction traveltime tomography (Prieux et al., 2009). The deep part of the starting model was clearly smoother than a velocity model inferred from reflection traveltime tomography, as the one shown in Sirgue et al. (2009; their Figure 3).

Ten iterations were performed per damping factor, leading to 30 iterations per frequency inversion. V_P and V_S are the reconstructed parameters. The density is constant and assumed to be known in the inversion.

Two tests were performed with and without outliers in the data. For both tests, random uniform white noise was introduced into the observed data and computed using the forward-problem engine implemented in the inversion code (the so-called inverse crime). The observed data were computed using a (Dirac) delta function for the

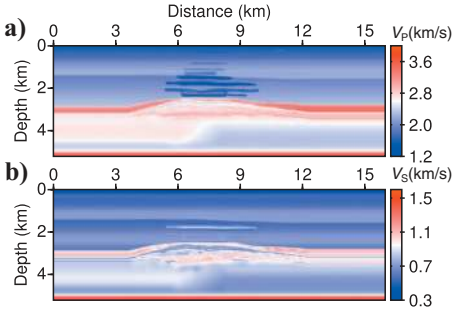

Figure 2. The true synthetic Valhall model for (a) P-wave and (b) S-wave velocities.

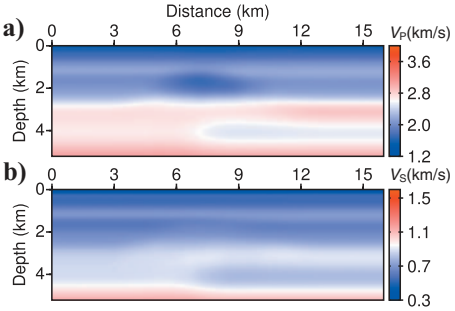

Figure 3. (a) V_P and (b) V_S starting models for FWI, as inferred by Gaussian smoothing of the true models (Figure 2).

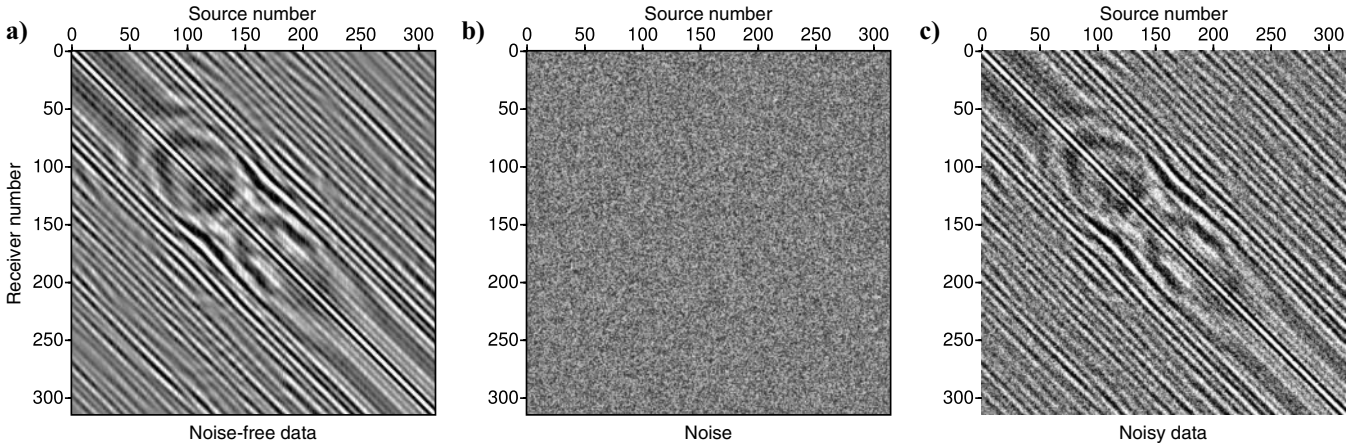

Figure 4. Real part of the 4-Hz frequency-domain data in the source/receiver domain for the Valhall model. (a) Noise-free hydrophone data; (b) added noise; (c) resulting contaminated data used for FWI.

source wavelet. Therefore, each frequency component of the data had the same S/N because white noise was considered. However, the source wavelet spectrum had negligible influence in frequency-domain FWI, where single frequencies or groups of frequencies of narrow bandwith are generally inverted sequentially — independently — in the framework of multiscale approaches.

The S/N was set to 10 dB, based on the power value P of the signal defined for one frequency by

$$\mathrm{S/N} = 10\log_{10}\frac{P_{\mathrm{data}}}{P_{\mathrm{noise}}} \text{ with } P = \sum_{i\mathrm{shot}=1}^{n\mathrm{shot}}\sum_{i\mathrm{rec}=1}^{n\mathrm{rec}} d_{i\mathrm{shot},i\mathrm{rec}} d^{*}_{i\mathrm{shot},i\mathrm{rec}}. \quad (13)$$

Figure 4 shows the 4-Hz noise-free and noisy data in the source/receiver domain for the hydrophone data.

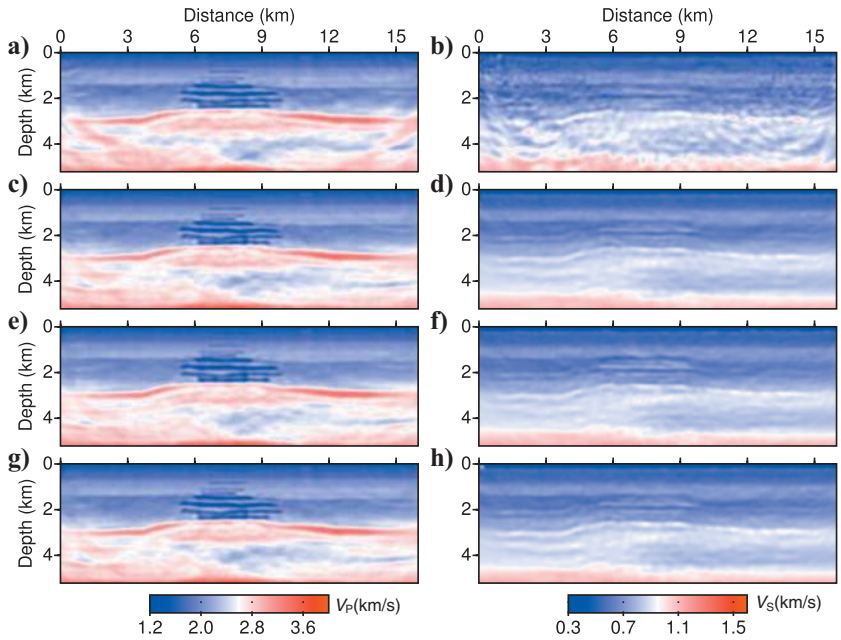

Figure 5. Reconstructed (left) V_P and (right) V_S models for the first Valhall test with the noisy data after the two FWI steps. (a, b) ℓ_2-norm; (c, d) ℓ_1-norm; (e, f) Huber criterion; (g, h) hybrid criterion.

Results

During the first test, we considered only the ambient noise. The V_P and V_S models inferred from the four minimization criteria (ℓ_2, ℓ_1, Huber, hybrid) after the second inversion step are shown in Figure 5. These reveal very good results for V_P models for all functionals, whereas only the robust ℓ_1-norm and the Huber and hybrid criteria provide acceptable V_S models.

In a second test, we introduced outliers into the data. Large errors (i.e., the noise was locally multiplied by 20) were introduced randomly in one trace out of 100 to simulate a poorly preprocessed data set. The resulting noise was consequently no longer uniform for this test. The V_P models obtained after the first inversion step with the four functionals are shown in Figure 6. The ℓ_1-norm and the Huber and hybrid criteria provide accurate V_P models, whereas the inversion rapidly converges toward a local minimum when the ℓ_2-norm is used. For the ℓ_2-norm, the inversion stops close to the first step because of the convergence toward a local minimum. The results obtained after the second inversion step performed with the ℓ_1-norm reliably reconstruct the V_P and V_S models (Figure 7), which are close to those obtained from data without outliers (Figure 5c and d). This highlights the limited sensitivity of the ℓ_1-norm to outliers even for V_S reconstruction.

NUMERICAL TESTS: ONSHORE SEG/EAGE OVERTHRUST MODEL

Inversion setup

A second numerical example focuses on the SEG/EAGE overthrust model. High-amplitude surface waves are present in the data of this onshore model and need to be taken into account during inver-

sion. Brossier et al. (2009a) design a hierarchical scheme to invert the body and surface waves jointly by using simultaneous frequency inversion. We followed this approach. The acquisition geometry was composed of 199 explosive sources 20 m below the surface; 198 vertical and horizontal geophones recording wavefields were located on the free surface.

Five discrete frequencies, distributed among two slightly overlapping frequency groups, were inverted: (1.7, 2.0, 3.5) and (3.5,4.8,7.2) Hz. The choice of these two groups of frequencies was shown to be efficient for inversion while limiting the computational cost (Brossier et al., 2009a). Five time-damping factors were used in cascade for each frequency group ($\gamma = 1.5,1.0,0.5,0.1,0.033$). We incorporated more damping factors and a smaller interval between damping factors than for the Valhall example because the overthrust case study is more nonlinear than the Valhall one and hence requires injecting more progressively increasing amounts of data. This increasing nonlinearity results from the presence of high-amplitude surface waves in the onshore overthrust case study.

Forty-five nonlinear iterations were performed per damping factor. Figure 8 shows the true V_P model of the overthrust and the 500-m Gaussian smoothed version used as the starting model. A constant Poisson's ratio of 0.24 was fixed to build the true and starting V_S models. The density was constant and assumed to be known during the inversion. V_P and V_S are the reconstructed parameters for the inversion of the horizontal and vertical components of the particle velocity.

The inverted data were computed with a Dirac source wavelet. Random uniform white noise was introduced into the observed data, with an S/N of 7 dB for each frequency component. Figure 9 shows the 3.5-Hz noise-free and noisy data in the source/receiver domain for the horizontal component of particle velocity.

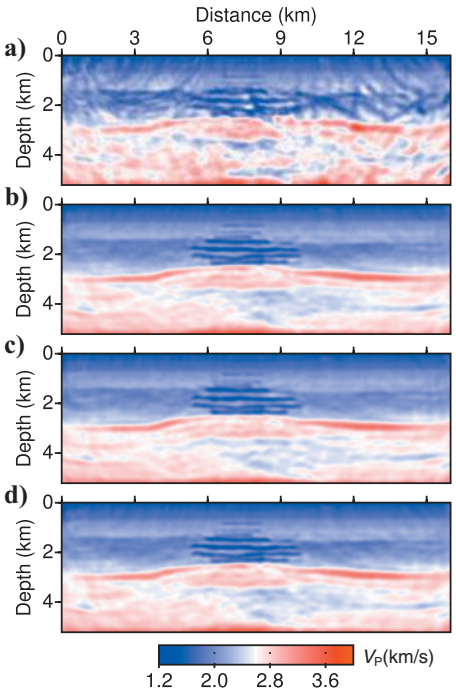

Figure 6. Reconstructed V_P models for the second Valhall test with the noisy data containing outliers, after the first FWI step. (a) ℓ_2-norm; (b) ℓ_1-norm; (c) Huber criterion; (d) hybrid criterion.

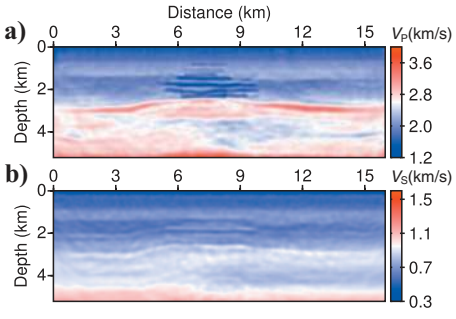

Figure 7. Reconstructed (a) V_P and (b) V_S models for the second Valhall test with the noisy data containing outliers after the two FWI steps with the ℓ_1-norm. The models are very close to those of Figure 5c and d.

Results

The V_P and V_S models obtained with the different minimization criteria are shown in Figure 10. The ℓ_1-norm again provides the most reliable results. The Huber and hybrid criteria show quite robust results, particularly in the shallow part of the models, even if the hybrid criterion suffers from high-frequency artifacts despite the smoothing regularization operator applied to the perturbation model at each iteration. The models obtained with the ℓ_2-norm are polluted by strong artifacts, particularly in the thrust area and in the deep structure.

DISCUSSION

Offshore Valhall model

The results of the first test, where only ambient noise without outliers was considered, show reliable reconstruction of the V_P model for the four norms, whereas only the ℓ_1, Huber, and hybrid functionals reliably reconstruct the V_S model.

In this shallow-water environment with low-velocity contrasts at the seabed, V_P imaging is more linear than V_S imaging for two main reasons. First, the larger P-wavelengths are resolved less well than their S-wave counterparts and are therefore less sensitive to the inaccuracies of the starting model in the framework of a multiscale reconstruction (Brossier et al., 2009a). Second, the P-waves dominate the seismic wavefield, whereas the P-to-S waves have a weaker foot-

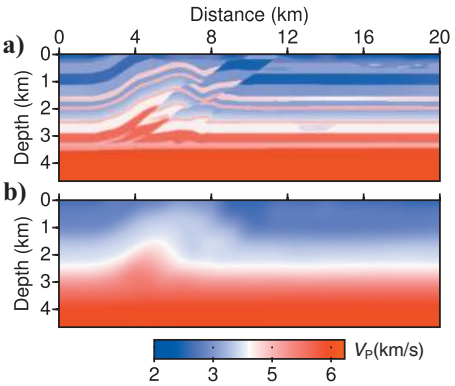

Figure 8. (a) True and (b) starting V_P models for the synthetic SEG/EAGE overthrust tests. The V_S models are derived from the V_P models using a constant Poisson ratio of 0.24.

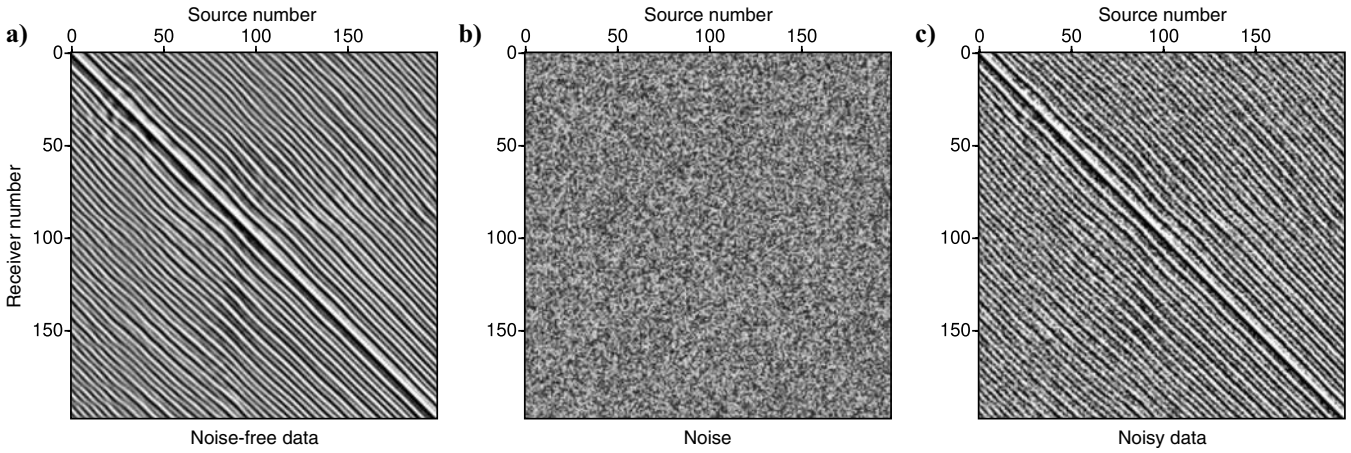

Figure 9. Real part of the 3.5-Hz frequency-domain data in the source/receiver domain for the overthrust model. (a) Noise-free horizontal component of particle velocity recorded by geophones; (b) added noise; (c) resulting contaminated data used for FWI.

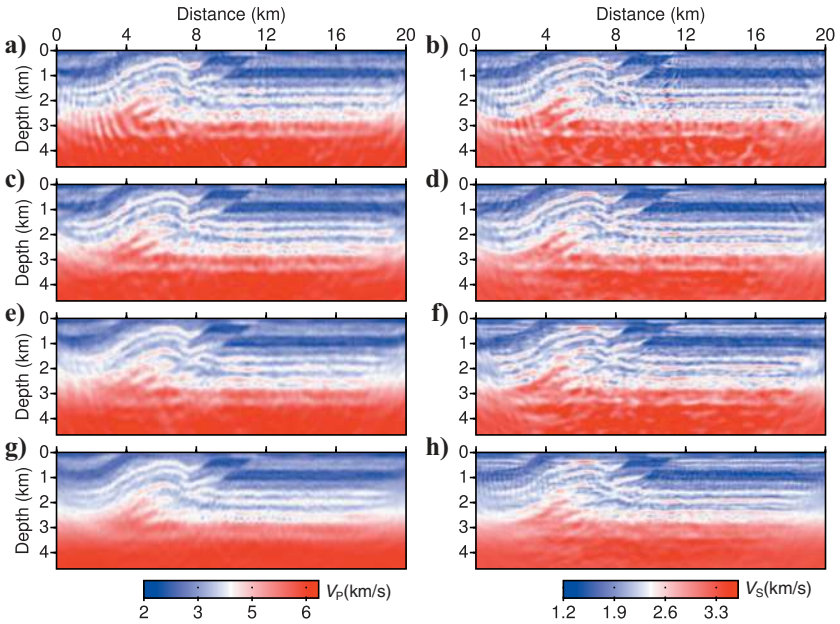

Figure 10. Reconstructed (left panels) V_P and (right panels) V_S models for the overthrust test obtained by FWI. (a–b) ℓ_2-norm; (c–d) ℓ_1-norm; (e–f) Huber criterion; (g–h) hybrid criterion.

print in the data. The limited signature of the S-waves in the data makes the inversion poorly conditioned for the S-wave parameter class, even with noise-free data.

Brossier et al. (2009b) show how the hierarchical two-step strategy allows us to increase the sensitivity of the inversion to the V_S parameter during the second step and, hence, the successful reconstruction of the V_S model with the ℓ_2-norm in the case of noise-free data. However, adding noise to the data still contributes to the weakened sensitivity of the FWI to the P-to-S arrivals. In this case, the two-step strategy implemented with the ℓ_2-norm failed to reconstruct the V_S model. In contrast, the ℓ_1-norm and the Huber and hybrid criteria successfully converged toward acceptable V_S models by mitigating the contribution of the residual amplitude in the reconstruction.

The second test, where outliers were introduced into the data, illustrates the expected failure of the ℓ_2-norm in the presence of high-amplitude isolated noise. The ℓ_2-norm intrinsically amplifies the weight of the high-amplitude residuals during inversion, causing divergence of the optimization if the residuals do not correspond to useful seismic arrivals. The ℓ_1-norm as well as the Huber and hybrid criteria show stable behavior for V_P imaging in this unfavorable context because the isolated, high-amplitude outliers have a negligible contribution in these functionals. The strong robustness of the ℓ_1-norm with respect to noise is illustrated by its ability to reconstruct the V_S model from low-amplitude P-to-S converted waves, even in the presence of outliers.

Onshore SEG/EAGE overthrust model

In an onshore context where body waves and surface waves are jointly inverted, data are very sensitive to V_P and V_S parameters. For example, V_S velocities on the near surface have a significant impact on high-amplitude surface waves. If the starting V_P and V_S models for FWI are not sufficiently accurate, then high-amplitude surface-wave residuals can direct the inversion toward a local minimum of the misfit function. In this context, data redundancy provided by the simultaneous inversion of multiple frequencies is essential to converge toward acceptable models (Brossier et al., 2009a). In the case of noisy data, however, it might be necessary to strengthen this redundancy by decreasing the frequency interval within each frequency group when the ℓ_2-norm is used.

Figure 11 shows the V_P and V_S models obtained with the ℓ_2-norm when the number of frequencies per group increases from three to nine: (1.7, 1.8, 2.0, 2.3, 2.5, 2.7, 3.0, 3.2, 3.5) and (3.5, 3.8, 4.1, 4.4, 4.8, 5.2, 6.0, 6.5, 7.2) Hz. Improvements of the V_P and V_S models, compared to those of Figure 10a and b, clearly show that increasing the data redundancy improves the S/N of the models at the expense of computing cost. On the contrary, the ℓ_1-norm shows more stable behavior in the presence of noise and therefore provides

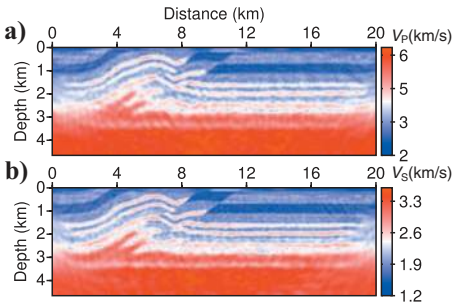

Figure 11. Reconstructed (a) V_P and (b) V_S overthrust models after FWI using the ℓ_2-norm applied to a dense data set of two slightly overlapping groups of nine frequencies each. Note the improvement in the S/N compared to Figure 10a and b.

more reliable results with efficient frequency-domain FWI, where only a few frequencies are involved in the inversion. The Huber and hybrid criteria show less convincing results than the Valhall experiment, highlighting the difficult tuning of the threshold ϵ.

Implications for 3D FWI

Some implications for 3D FWI can be derived from the results of the synthetic experiments. Historically, the two main motivations behind 2D frequency-domain modeling and inversion are (1) to design computationally efficient algorithms by limiting modeling and inversion to a few discrete frequencies and (2) to design multiscale strategies by proceeding from low frequencies to higher ones (Pratt and Worthington, 1990). For 3D problems, comparative analyses of modeling methods has shown that time-domain explicit-scheme modeling provides a computationally efficient alternative to frequency-domain modeling methods based on direct or iterative solvers to perform frequency-domain FWI (Nihei and Li, 2007; Plessix, 2007; Sirgue et al., 2008; Virieux et al., 2009).

Our results have highlighted the benefits provided by data redundancy for performing robust elastic FWI, where V_P and V_S are reconstructed and the ℓ_2-norm is used. Therefore, is it worth leaving the inversion in the frequency domain if the time-domain modeling engine provides the opportunity to exploit the data redundancy fully by means of time-domain FWI? A more quantitative analysis of the computational burden that results from implementing the inversion in the time domain compared to the frequency domain might be needed to answer this question. The computational burden resulting from implementing the inversion in the time domain might result from disk access or extra forward-problem resolutions that depend on the implemented numerical strategy (Akcelik, 2002; Liu and Tromp, 2006; Symes, 2007). Multiscale strategies in time-domain FWI can be implemented using the approach of Bunks et al. (1995), where successive inversions of data sets of increasing high-frequency content are performed. If efficient 3D elastic frequency-domain FWI must be performed, the ℓ_1-norm definitively provides the more robust criterion for the noise levels investigated in this study.

CONCLUSIONS

Application of elastic FWI to offshore and onshore data shows the strong sensitivity of the ℓ_2-norm to non-Gaussian errors in the data when decimated discrete frequencies in FWI are considered for computational efficiency. The marginally used ℓ_1-norm appears to be weakly sensitive to noise, even in the presence of outliers. It provides stable results for onshore and offshore FWI applications. In particular, the ℓ_1-norm allows successful inversion of low-amplitude P-to-S arrivals for reliable V_S model building in an offshore environment, even in the presence of noise. Alternative functionals such as the Huber and hybrid criteria, which combine the ℓ_1- and ℓ_2-norms, can provide stable results if the threshold controlling the transition between the ℓ_1 and the ℓ_2 behaviors is well chosen. The judicious estimation of this threshold by trial-and-error tests is a clear drawback of the Huber and hybrid criteria, even if these functionals should be as good as the ℓ_1-norm if the threshold is well chosen. More automatic functionals, such as the ℓ_1- or ℓ_2-norms, should therefore be recommended for inversion of field data.

The ℓ_1-norm reveals an interesting alternative to the ℓ_2-norm, especially when decimated data sets are processed by efficient frequency-domain FWI. However, the results obtained with the ℓ_2-norm can be improved in the presence of ambient noise by increasing the data redundancy through simultaneous inversion of multiple frequencies at the expense of computing costs.

Some implications for 3D FWI can be derived from these conclusions. On the one hand, the ℓ_1-norm might be an appealing alternative to the ℓ_2-norm for FWI based on frequency-domain seismic modeling, which requires consideration of a limited number of frequencies for computational efficiency. On the other hand, explicit time-marching algorithms are competitive with frequency-domain seismic modeling methods for 3D problems. If the wavefields are computed fully in the time domain for computational efficiency, the inversion of decimated data in the frequency domain may not provide a significant computational savings compared to a time-domain inversion. In this case, the inversion may be left in the time domain to take full advantage of data redundancy.

ACKNOWLEDGMENTS

This study was funded by the SEISCOPE consortium, sponsored by BP, CGGVeritas, ExxonMobil, Shell, and TOTAL and by Agence Nationale de la Recherche (ANR) under project ANR-05-NT05-2-42427. The LU factorization of the impedance matrix was performed with MUMPS software (http://graal.ens-lyon.fr/MUMPS/index.html). The mesh generation was performed with the help of Triangle software (http://www.cs.cmu.edu/~quake/triangle.html). Access to the high-performance computing facilities of the MESOCENTRE SIGAMM computer center and IDRIS national (project 082280) centers provided the required computer resources; we gratefully acknowledge these facilities and the support of their staff. Our thanks go to J. Kommendal and L. Sirgue from BP for providing the elastic synthetic models of Valhall. We would like to thank Bill Harlan, Partha Routh, and three anonymous reviewers for their fruitful comments on the manuscript.

REFERENCES

Akcelik, V., 2002, Multiscale Newton-Krylov methods for inverse acoustic wave propagation: Ph.D. dissertation, Carnegie Mellon University.

Amestoy, P. R., A. Guermouche, J. Y. L'Excellent, and S. Pralet, 2006, Hybrid scheduling for the parallel solution of linear systems: Parallel Computing, **32**, 136–156.

Brenders, A. J., and R. G. Pratt,2007a, Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.

———, 2007b, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.

Brenders, A., R. Pratt, and S. Charles, 2009, Waveform tomography of 2-d

seismic data in the Canadian foothills — Data preconditioning by exponential time-damping: 71st Conference & Exhibition, EAGE, Extended Abstracts, U041.
Brossier, R., S. Operto, and J. Virieux, 2009a, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, no. 6, WCC63–WCC76.
———, 2009b, Two-dimensional seismic imaging of the Valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion: 79th Annual International Meeting, SEG, Expanded Abstracts, 2293–2297.
Brossier, R., J. Virieux, and S. Operto, 2008, Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling: Geophysical Journal International, **175**, 541–559.
Bube, K. P., and R. T. Langan, 1997, Hybrid l_1/l_2 minimization with applications to tomography: Geophysics, **62**, 1183–1195.
Bube, K. P., and T. Nemeth, 2007, Fast line searches for the robust solution of linear systems in the hybrid l_1/l_2 and Huber norms: Geophysics, **72**, no. 2, A13–A17.
Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.
Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.
Djikpéssé, H. A., and A. Tarantola, 1999, Multiparameter l_1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035.
Etienne, V., J. Virieux, N. Glinsky, and S. Operto, 2009, Seismic modelling with discontinuous Galerkin finite-element method — Application to large scale 3D elastic media: 71st Conference & Exhibition, EAGE, Extended Abstracts, P131.
Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the Huber norm: Geophysics, **68**, 1310–1319.
Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: Geophysics, **74**, no. 3, R15–R24.
Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: The Annals of Statistics, **1**, 799–821.
Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: Geophysical Journal International, **157**, 1269–1296.
Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator: Geophysics, **61**, 529–537.
Kommedal, J. H., O. I. Barkved, and D. J. Howe, 2004, Initial experience operating a permanent 4C seabed array for reservoir monitoring at Valhall: 74th Annual International Meeting, SEG, Expanded Abstracts, 2239–2242.
Liu, Q., and J. Tromp, 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seismological Society of America, **96**, 2383–2397.
Munns, J. W., 1985, The Valhall field: A geological overview: Marine and Petroleum Geology, **2**, 23–43.
Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD): Geophysical Journal International, **169**, 1069–1078.
Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.
Operto, S., J. Virieux, P. Amestoy, J.-Y. L'Excellent, L. Giraud, and H. Ben-Hadj-Ali, 2007, 3D finite-difference frequency-domain modeling of viscoacoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, no. 5, SM195–SM211.
Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.
———, 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185–SM194.
———, 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, no. 6, WCC53–WCC61.
Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method: Geophysical Prospecting, **38**, 311–330.
———, 1999, Seismic waveform inversion in the frequency domain, part I — Theory and verification in a physic scale model: Geophysics, **64**, 888–901.
Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.
Prieux, V., S. Operto, R. Brossier, and J. Virieux, 2009, Application of acoustic full waveform inversion to the synthetic Valhall model: 79th Annual International Meeting, SEG, Expanded Abstracts, 2268–2272.
Pyun, S., C. Shin, and W. Son, 2009, Frequency-domain waveform inversion using an L1-norm objective function: 71st Conference & Exhibition, EAGE, Extended Abstracts, P005.
Sears, T., S. Singh, and P. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.
Shin, C., S. Jang, and D. J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.
Shin, C., D.-J. Min, K. J. Marfurt, H. Y. Lim, D. Yang, Y. Cha, S. Ko, K. Yoon, T. Ha, and S. Hong, 2002, Traveltime and amplitude calculations using the damped wave solution: Geophysics, **67**, 1637–1647.
Sirgue, L., O. I. Barkved, J. P. V. Gestel, O. J. Askim, and J. H. Kommedal, 2009, 3D waveform inversion on Valhall wide-azimuth OBC: 71st Conference & Exhibition, EAGE, Extended Abstracts, U038.
Sirgue, L., J. T. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Conference & Exhibition, EAGE, Extended Abstracts, F022.
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.
Sourbier, F., A. Haidar, L. Giraud, S. Operto, and J. Virieux, 2008, Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D full-waveform inversion?: 78th Annual International Meeting, SEG, Expanded Abstracts, 2147–2151.
Stekl, I., and R. G. Pratt, 1998, Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators: Geophysics, **63**, 1779–1794.
Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, no. 5, SM213–SM221.
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.
———, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Scientific Publ. Co., Inc.
Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least square criterion: Reviews of Geophysics and Space Physics, **20**, 219–232.
Virieux, J., S. Operto, H. Ben-Hadj-Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar, 2009, Seismic wave modeling for seismic imaging: The Leading Edge, **28**, 538–544.
Warner, M., I. Stekl, and A. Umpleby, 2007, Full wavefield seismic tomography — Iterative forward modelling in 3D: 69th Conference & Exhibition, EAGE, Extended Abstracts, C025.